



## Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

# THE JOURNAL OF PHILOSOPHY

---

## EMPIRICISM VERSUS FORMALISM IN WORK WITH MENTAL TESTS

THE past four or five years have seen a simply extraordinary activity in the field of "mental tests." "Testing" has, however, developed in a curious fashion largely without definite scientific affiliations. As one result, critical discussion of fundamental principles has lagged far behind application. And work in this field is tied up with large hypotheses which have never received adequate critical consideration. As one of a number of recent efforts toward a critical review of methods, the writer has lately published a paper entitled—somewhat ambitiously—"Suggestions Looking Toward a Fundamental Revision of Current Statistical Procedure, as Applied to Tests."<sup>1</sup> He has just received from Dr. Ruml a proof of a paper commenting on this article.<sup>2</sup> With the proof was a letter in which Dr. Ruml expresses the hope that a reply to these comments may be early forthcoming; he feels that discussion upon the mental principles in test work is at present most desirable. With this last statement, at least, the writer is in full agreement; and he hastens to make such reply. The issues can, he believes, be put very briefly. They are three in number.

1. The fundamental contention of the writer is for a more empirical method, in work with tests, and less easy assumption of large hypotheses. For instance, it should be obvious that if the validity of the theory of general intelligence is assumed, and results on different tests summed to make a total score, if teachers or employers are asked to estimate "general ability," if marks in school or college are averaged to give general ability in school work, little evidence contrary to the theory of general ability can be found. The whole situation has been generalized in the first place. The concept of general intelligence is, as a matter of fact, being very seriously questioned at present.<sup>3</sup> The contribution, to this fundamental

<sup>1</sup> *Psychological Review*, Vol. 27, pp. 466, 472, November, 1920.

<sup>2</sup> Beardsley Ruml, "Reconstruction in Mental Tests," this JOURNAL, Vol. 18, pp. 181-185, March 31, 1921.

<sup>3</sup> For an admirable non-technical account of some recent conclusions in the matter, see E. L. Thorndike, "Measuring Human Intelligence," *Harper's Magazine*, Vol. 140, 1920, pp. 227-235. The same writer summarizes some interesting correlations in a brief paper "Reliability and Significance of Tests of Intelligence," *Jour. of Educational Psychology*, Vol. 11, 1920, pp. 284-87. An interesting discussion from the point of view of the employment psychol-

problem, of the tremendous amount of work with mental tests is really astoundingly small, simply because the work has been so tied up with the hypothesis.

The same situation holds with regard to the use of the theory of the "normal" curve, in test work. It is common practise, in building educational tests, to assume that the ability in question is distributed according to the "normal" curve. The items are, then, selected in the first place to give equal units on a "probable error" scale. And, naturally, the results given with the test show a "normal" distribution. The investigator then goes further (very likely) and works out supposedly equal units of ability, calculates a "zero point" in the ability in question, as measured by the scale, and talks easily about "silent reading ability" or general intelligence, in the same fashion that he might discuss measures of height or weight. As a matter of fact, the concrete facts of the situation are so overlaid by the theory which has been involved in the treatment that the validity of the theory for the particular ability in question can not be determined, nor the application of the theory to the practical problem concerned.<sup>4</sup>

Other large assumptions, perhaps less vital, but no less unfortunate in principle, are common in present statistical practise. The involvement of the test work in such hypotheses is surely unfortunate. It is also decidedly unfortunate that such methods have led to the development of a formal statistical method; and we find effort that should be devoted to the study of fundamental problems being wasted, instead, in the fulfilment of formal statistical requirements. An astounding amount of time is being devoted to the development of "units," "zero points," and "weightings"—to the mystification of the uninitiated, and to the neglect of elementary problems of analysis and verification.

When a situation is thus clouded by hypotheses, a return to a rigidly empirical method is naturally suggested. A mode of attack must be worked out which will be free from such doubtful hypothetical elements. So the writer has argued (to Dr. Ruml's horror, apparently) for, "no hypotheses, as thoroughly empirical treatment as may be!" Surely such a return to rigidly empirical

ogist will be found in Henry C. Link's *Employment Psychology*, Macmillan, 1919, Chapter 12, "General Intelligence, a Dialogue," pp. 130-139. Consideration of theoretical aspects of the matter will be found in articles by G. H. Thompson, "The General Factor Fallacy in Psychology," *British Journal of Psychology*, Vol. 10, 1920, pp. 319-26.

<sup>4</sup> See, for instance, E. G. Boring, "The Logic of the Normal Law of Error in Mental Measurements," *American Journal of Psychology*, Volume 31, 1920, pp. 1-33. The paper is, the writer believes, of the very greatest theoretical importance.

methods would seem the natural and healthy way out of such a situation. It would seem the only way in which present hypotheses might be verified, and new and more sound hypotheses developed.<sup>6</sup>

2. The writer's second point has to do simply with a matter of fact. *As a matter of actual fact*, "testing" is at present largely a technical science. The writer has argued, as strongly as any one, for purer research in the field of mental tests and for the need for such research.<sup>7</sup> He is now devoting more time than he should, perhaps, to such work. However, *as a matter of actual fact*, work with tests is now largely of an immediately practical nature—or at least, is with reference to practical problems. Under such circumstances, certain very definite and very practical obligations are incurred. Suppose one is using tests of intelligence in order to determine whether a boy in the fourth grade may not be capable of doing fifth grade work; the hope is, that by a little special coaching and "double-promotion" the child may ultimately be saved a year in his school work. It is no abstract child who is under discussion—it might be your child. The one consideration which is paramount to all others, under such circumstances, is the very practical consideration—which tests, and which methods, will aid most effectively in deciding this practical question. Dr. Ruml comments, in horrified tones, on the fact that the writer "would even use his statistics to produce bimodal distributions," in order to deal most efficiently with such a practical problem. Certainly the writer would be quite willing to sacrifice formal statistical methods in order to deal more effectively with the boy.

The contrary practise is, unfortunately, the common practise at present. The one great question would seem to be: just what will these tests tell with regard to most satisfactory grade placement? How well (the question is) will the test inform us as to whether this boy can do the work of the fifth grade, and what is the significance of a score of 20, in this matter, as compared with a score of 15? Instead, investigators are putting their best efforts into finding out whether a score of 20 is as far above 15 as 15 is above 10, and how far these various scores are above "zero" ability.

<sup>6</sup> Ruml issues a solemn warning against any such "direct action." The empirical method as "direct action"! Such strictures have, the writer believes, been put upon efforts at a return to first facts, before.

<sup>7</sup> See especially S. L. and L. W. Pressey, "Cross-out Tests, with Suggestions as to a Group Scale of the Emotions," *Journal of Applied Psychology*, Vol. III, 1919, pp. 138-150. Also, S. L. Pressey and O. R. Chambers, "First Revision of a Group Scale Designed for Investigating the Emotions with Tentative Norms," *Journal of Applied Psychology*, Vol. IV, 1920, pp. 97-104; and L. W. and S. L. Pressey, "A Critical Study of the Concept of Silent Reading Ability," *Journal of Educational Psychology*, Vol. XII, January, 1921, pp. 25-31.

It is such substitution of formal statistical requirements for practical requirements, in use of tests for dealing with practical problems, to which the writer particularly objects. Practical efficiency is certainly the criterion by which our statistical technique should be judged, if we are dealing with practical problems. Our obligation is not to the requirements of a formal method, but to the needs of the people with whom we are working. If a "statistical trick" will inform us more accurately as to whether the boy above mentioned belongs in the fifth grade, then by all means use the "trick." Instead, it more often happens that the best welfare of the boy is sacrificed to the requirements of statistical pedantry.

To sum up then: the writer wishes (1) a return to strictly empirical methods in view of current injudicious and uncritical use of hypotheses, and (2) a recognition of the practical obligations involved in use of tests for practical purposes, and the development of statistical methods designed specifically with reference to these practical problems. Now, a word with regard to the type of method which, to the writer's view, should give the empirical approach desired, and as to the relation of such a method to testing as a technical science. It must be emphasized that any faults of this method do not necessarily comprise the truth of the first two contentions; the writer insists that more empirical method and a more consciously technical treatment for technical problems, are desirable; any faults of his particular type of empiricism, or of his notions with regard to the relations of empiricism to a psycho-technique, must not be allowed to compromise the main issues. However, the first two contentions may be made clearer if some positive suggestions as to method are added.

3. It is first necessary, as the writer sees it, that methods be freed from involvement in hypotheses as yet not thoroughly verified (a requirement that certainly leaves very few hypotheses which may safely be thus employed!) To return to the previous example, it should not be assumed that the same tests may be used in prognosticating success or failure in learning to read, and in learning arithmetic. Each problem must be dealt with separately. If it appears that the same tests may be used to indicate capacity along these two lines, well and good. But this must not be assumed. There must be specific study of specific problems, and always an opportunity for analysis.

There must also (in such empirical procedure) be a willingness to work with materials as they are, without transformation into terms more readily treated according to current statistical practices. To return again to a previous example, it may be that "arithmetical ability" is distributed according to the "normal"

curve. It may well be that problem-solving is at present taught in the school in such a haphazard way that the operation of chance factors does bring about a "chance" distribution. But one must not proceed to do all his thinking with regard to the teaching of arithmetic in these terms, from now on. The aim, in the teaching of problem-solving, is to equip the child to solve common practical problems in later life; and there is no evidence that such practical requirements are so distributed. In fact, the writer suspects, it is a fairly reasonable inference that the most efficient school would bring about a distribution which would not be "normal." Moreover, schools are organized at present primarily with reference to minimal requirements. And the fundamental unit is the grade. An empirical treatment will seek to construe the data with reference to these various facts, rather than in terms of a biological or statistical theory.

And here we come (if the writer understands Ruml correctly) to the third point of difference between Dr. Ruml and himself. Such empirical methods have the great advantage of also getting closer to practical problems. In other words, the writer believes that the study of practical problems will not lead one to "ignore the necessity for analysis," but will bring about analysis; will not "warp the observed data," but will describe the data as they are. Such methods will "warp" the results into a shape not congruous with some current theories, to be sure. But that is not proof positive that these methods are wrong. One can not but wonder whether Dr. Ruml's difficulty may not be that he is thinking entirely in terms of "Pearsonian statistics" while we need a "relativistic" theory.... Anyhow, the writer refuses to be frightened away from the greatest possible efficiency in his technical work, because of the charge that these most efficient methods are unscientific. Rather, he believes that the efficiency of a method is not such bad evidence of its fundamental soundness. And he will be inclined to take his cue from this, and aim first at the greatest usefulness, in test methods—with the confidence that such a criterion could not lead one very far astray.

Well—so much for the general concept, as the writer sees it. However, it is obvious that his theory, essentially, can not rest with a general statement but calls for a pragmatic proof. The question is as to what, actually, such an approach will yield. He will have in print, shortly, test materials developed and handled precisely after the method suggested in the previous paper.<sup>8</sup> It is

<sup>8</sup> In fact, he has already presented something of this sort in previous papers. See, for instance, S. L. and L. W. Pressey, "Irregularity in a Psychological

about such concrete matter rather than about general statements of theory that discussion would seem most profitably centered. That there should be such discussion seems to the writer most highly desirable; he is in thorough agreement with Dr. Ruml on this point. In fact, he hopes, with Dr. Ruml, that others may take part in the discussion. He also hopes, with him, that any such discussion may proceed without any of the personal irritation which sometimes so unhappily develops, in the course of differences regarding matters of scientific method. There is surely a need for the clarification of both aims and methods in the field of tests. Such clarification is most decidedly needed, if test work is to continue its healthy growth. If "testing" can thus be brought into its own, the writer believes, as stated in a previous article<sup>9</sup> that the test method will be of great value as a method in pure research. Meanwhile, testing is primarily, at present, of a practical nature; and the writer feels that the most healthy and consecutive development, from both the scientific and practical point of view, can most economically be obtained by a vigorous, persistent, and open-minded carrying through of the practical problems to which we are now obligated.

S. L. PRESSEY.

INDIANA UNIVERSITY.

---

## A NOTE ON THE RÔLE OF MATHEMATICS IN PHYSICS

### I

BECAUSE *static* phenomena furnished the model upon which Cartesian geometry was constructed, Leibniz and Newton, especially the latter, were compelled to develop or use a mathematical theory built upon the model of *dynamic* facts as supplied primarily by Da Vinci, Copernicus, Galileo and Kepler. If it is fertile, calculus, as well as any other system of mathematical theory, should bring to light new facts about the physical phenomena as a result of which it was originally constructed. Furthermore the new facts give rise to a new model of physical phenomena. In fact, every new physical fact gives birth to a new model of physical phenomena, upon which mathematicians develop new mathematical theories, which in turn give rise to new generalizations, and if taken up by physicists are the means of discovering new existential facts.

Examination as a Measure of Mental Deterioration," *Journal of Abnormal Psychology*, Vol. XIII, No. 5, December, 1918, pp. 285-294; and S. L. and L. W. Pressey, "The Practical 'Efficiency' of a Group Scale of Intelligence," *Journal of Applied Psychology*, Vol. III, March, 1919, pp. 68-80.

<sup>9</sup> See note 7.